JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 76, 483–497 (1980)

Randomness Implies Order*

ABRAHAM BOYARSKY

Department of Mathematics, Sir George Williams Campus, Concordia University, Montreal, Canada

Submitted by G.-C. Rota

Let τ: [0, 1] $\to$ [0, 1] possess a unique invariant density f^*. Then given any $\epsilon > 0$, we can find a density function p such that $\| p - f^* \| < \epsilon$, and p is the invariant density of the stochastic difference equation $x_{n+1} = \tau(x_n) + W$, where W is a random variable. It follows that for *all* starting points $x_0 \in [0, 1]$, $\lim_{n\to\infty} (1/n) \sum_{i=0}^{n-1} \chi_B(x_i) = \int_B p(\xi)\, d\xi$.

INTRODUCTION

There are many transformations which possess an infinite number of periodic points. In the important paper [14], it was shown that if a continuous transfor mation τ from an interval of the real line into itself has a cycle of period three, then it must have cycles of all orders. As well, there exists an uncountable set of points $\mathscr{D}$ such that an orbit starting in $\mathscr{D}$ does not approach any cyclic point. The combination of these properties is referred to as "chaos." Since the chaotic functions are dense in the space of continuous functions [15], there are many such chaotic functions.

From the point of view of statistical analysis, the real chaos is often contributed by the periodic points. Whereas for orbits starting in $\mathscr{D}$ we can hope for uniform long-term statistical behaviour, this is clearly impossible for the periodic points. An experimental researcher may be more annoyed by this inability to reproduce cyclic results due to slight variations in the initial paramters than by the existence of a set of starting points for which the orbits behave erratically. One of the purposes of this note is to present a technique for overcoming the chaos engendered in difference equations by periodic points.

Let $J = [0, 1]$ and let τ: $J \to J$ have a unique absolutely continuous invariant measure μ with density f^*. The Birkhoff Ergodic Theorem states that f^* can be found by taking appropriate time averages; more generally, for g any meaningful measurement,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\tau^n(x)) = \int_J g(y) f^*(y)\, dy, \qquad \text{a.e. } \mu, \tag{1}$$

* This research was supported by a grant from the National Research Council of Canada, Grant A-9072.

0022-247X/80/080483-15$02.00/0

where $\tau^n \equiv \tau \circ \tau^{n-1}$ is the nth iterate of τ. In many problems, the exceptional set of μ-measure 0 may be prohibitively large. For example, for a large class of piecewise-linear transformations τ, the rationals are eventually periodic [1]. But the rationals are the only points with which computations can be performed. Thus, for such transformations, f^* cannot be found in practice by direct iteration of the difference equation

$$x_{n+1} = \tau(x_n). \tag{2}$$

Motivated by computational difficulties in implementing the Birkhoff Ergodic Theorem, Li [2] presented a technique, originally conjectured by Ulam [3], for approximating f^*. The method involves the construction of a sequence of "approximating" Frobenius–Perron operators, which, when restricted to the set of piecewise-constant functions, can be represented by larger and larger matrices. There are two drawbacks with this scheme: (a) the calculation of the fixed points of the matrices may be inaccurate and costly for transformations which are not piecewise linear, and (b) there is no mention of an underlying system, such as (2), which can be viewed as producing the approximating sequence of densities.

To evade cyclic orbits and to make the model (2) more realistic, we introduce a random perturbation term. The difference equation (2) then becomes a stochastic difference equation

$$x_{n+1} = \tau(x_n) + W_\lambda, \tag{3}$$

where for each $0 < \lambda < 1$, W_λ is a random variable having the probability density function ϕ_λ. We shall assume that as $\lambda \to 1^-$, $\phi_\lambda \Rightarrow \delta_0$, the point measure at $x = 0$, where $\Rightarrow$ denotes weak convergence. Since disturbances can rarely be avoided in nature, it is reasonable to study (3) rather than its deterministic, but often unpredictable counterpart (2). We say (2) is unpredictable because the limiting behaviour depends on the starting point. We shall see, however, that under general conditions, the stochastic difference equation (3) can be implemented for all starting points to obtain as close an approximation to μ as desired. In this sense, the random perturbation implies order: the unfailing acquisition of the (approximate) absolutely continuous invariant measure μ. We may regard this as a kind of structural stability of μ under stochastic perturbations.

2. Main Result

Let $(\mathscr{L}_1, \|\cdot\|)$ denote the space of all integrable functions defined on $J = [0, 1]$ and let $\tau: J \to J$ be a nonsingular transformation. The Frobenius–Perron operator $P_\tau: \mathscr{L}_1 \to \mathscr{L}_1$ is defined by

$$P_\tau f(x) = \frac{d}{dx} \int_{\tau^{-1}([0,x])} f(y)\, dy. \tag{4}$$

Let us suppose X is a random variable with probability density function $f(x)$, then the random variable $\tau(X)$ has a probability density function (by the Radon–Nikodym Theorem), and it is given by $P_\tau f$. Hence, we can interpret $P_\tau f$ as the density of $\tau(X)$.

For the moment, let us suppose that $\tau: R \to J$, and coasider the random variable $\tau(X_n) + W_\lambda$, where X_n is a random variable with density $f(x)$ and W_λ has probability density function $\phi_\lambda(x)$. If we assume that W_λ is independent of X_n, the density of $\tau(X_n) + W$ is $(P_\tau f) * \phi_\lambda$, where $*$ denotes convolution. Since we are interested in the limiting density of the equation

$$X_{n+1} = \tau(X_n) + W_\lambda, \tag{5}$$

we can assume that the density of X_{n+1} is also $f(x)$, and hence we require that

$$(P_\tau f) * \phi_\lambda = f. \tag{6}$$

The solution of the convolution equation (6) is an invariant density for the Markov chain (5). But what if the domain of τ is a bounded set? Then the random perturbation W_λ can take the process out of the domain at some iterate. Before dealing with this, let us define precisely the class of transformations τ with which we shall be concerned.

We say $\tau: J \to J$ is in $\mathscr{C}$ if (i) it is piecewise C^2, (ii) it satisfies $\inf_J |\tau'(x)| > 1$, and (iii) it possesses a unique absolutely continuous invariant measure. From (i) and (ii), it follows that there exists an absolutely continuous invariant measure under τ [6]. General conditions are known which ensure that the invariant measure is unique. From the results in [9], it follows that if τ and τ' have one (the same) point of discontinuity, then the absolutely continuous invariant density is unique. In [12], it is shown that a large class of Markov maps possess unique invariant measures. In [10], it is shown that if τ satisfies conditions (i) and (ii), then $p(x) = \tau(x) \pmod 1$ has a unique absolutely continuous invariant measures. Results for other classes of transformations are scattered in the literature.

Let the family of random variables $\{W_\lambda: 0 < \lambda < 1\}$ possess smooth probability density functions $\{\phi_\lambda(x): 0 < \lambda < 1\}$ with the properties that support $\phi_\lambda \subset [-a, a]$ for all λ and that

$$\phi_\lambda \Rightarrow \delta_0$$

as $\lambda \to 1^-$, where δ_x denotes the point measure at x and $\Rightarrow$ denotes weak convergence, namely

$$\int_{-a}^{a} h(x)\,\phi_\lambda(x)\,dx \to \int_{-a}^{a} h(x)\,\delta_0(x)\,dx = h(0)$$

for all continuous functions h on $[-a, a]$. The family $\{\phi_\lambda: 0 < \lambda < 1\}$ is called an "approximation of the identity."

Since W_λ can take values in $[-a, a]$, the right-hand side of (3) can take values in $J_a \equiv [-a, 1 + a]$. In order to use (3), we must extend τ to J_a. We do this as follows: define $\tau_a: J_a \to J$ by $\tau_a|_J = \tau$ and $\tau_a|_{J_a - J}$ is C^2 with $\inf |\tau_a'| > 1$ and $\tau_a(J_a - J) \subset \tau(J)$, i.e., the range of τ_a = range of τ. Then, replacing τ by τ_a in (3), the new stochastic difference equation

$$X_{n+1} = \tau_a(X_n) + W_\lambda \tag{3'}$$

is well defined, and has the property that if an orbit $\{\tau_a^i(x)\}_{i=0}^{\infty}$ is in J, then with respect to that orbit atleast (3′) reduces to (3). An invariant density of (3′) must satisfy $(P_{\tau_a} f) * \phi_\lambda = f$. Since τ_a is piecewise C^2 and satisfies $\inf_{J_a} |\tau_a'| > 1$ all the results of [6] apply to $P_{\tau_a} f$. We shall refer to τ_a as a $\mathscr{C}$-extension of τ; this transformation does not have to possess a unique invariant density.

We shall need a few preliminary results before proving the main theorem. Let $A = [-a, a]$ and let $\mathscr{M}(A)$ denote the set of probability density functions with support in A.

LEMMA 2.1. *Let $\{\phi_\lambda: 0 < \lambda < 1\} \subset \mathscr{M}(A)$ such that $\phi_\lambda \Rightarrow \delta_0$ as $\lambda \to 1^-$. Assume that each ϕ_λ is continuous. Then $\exists \{\psi_\lambda\} \subset \mathscr{M}(A)$, $\{\beta_\lambda\} \subset \mathscr{M}(A)$ such that*

$$\psi_\lambda = \lambda \delta_0 + (1 - \lambda) \beta_\lambda$$

and such that $\psi_\lambda - \phi_\lambda \Rightarrow 0$ as $\lambda \to 1^-$. Hence $\psi_\lambda \Rightarrow \delta_0$ as $\lambda \to 1^-$.

Proof. For each $\lambda_0 \leqslant \lambda < 1$, we can find $q_\lambda(x)$ such that $\phi_\lambda(x) \geqslant q_\lambda(x) \geqslant 0$ and

$$\int_A q_\lambda(x)\, dx = \lambda.$$

Set

$$\beta_\lambda = \frac{\phi_\lambda - q_\lambda}{1 - \lambda}.$$

Then $\beta_\lambda \in \mathscr{M}(A)$. Let

$$\psi_\lambda = \lambda \delta_0 + (1 - \lambda) \beta_\lambda .$$

Now write

$$\phi_\lambda = \lambda \delta_\lambda + (1 - \lambda) \left[\frac{\phi_\lambda - \lambda \delta_0}{1 - \lambda} \right].$$

Note that $(\phi_\lambda - \lambda\delta_0)/(1 - \lambda)$ is not a probability density function. Since $\phi_\lambda \Rightarrow \delta_0$, we must have $q_\lambda \Rightarrow \lambda\delta_0$. Hence,

$$\beta_\lambda \Rightarrow \frac{\phi_\lambda - \lambda\delta_0}{1 - \lambda}$$

as $\lambda \to 1^-$. Therefore, $\psi_\lambda \Rightarrow \phi_\lambda \Rightarrow \delta_0$ as $\lambda \to 1^-$.

EXAMPLE. Let ϕ_λ be the triangle with base $[1-\lambda,\ 1+\lambda]$ and height $1/(1-\lambda)$. Then q_λ can be chosen as the triangle with the same base and height $\lambda/(1-\lambda)$.

We shall denote the variation of a function f over the interval $[\alpha, \beta]$ by $\bigvee_\alpha^\beta f$ or $\bigvee_{[\alpha,\beta]} f$.

LEMMA 2.2. *Let f have support on $[b, c]$ and let it be of bounded variation. Let g have support on $[-a, a]$ and $\int_{-a}^{a} |g(t)|\, dt \leqslant 1$. Then*

$$\bigvee_{b-a}^{c+a} (f * g) \leqslant \bigvee_{b-a}^{c+a} f.$$

Proof. Note that $f * g$ has support in $[b-a, c+a]$. Let $\mathscr{P} = \{x_0, ..., x_n\}$ be a partition of $[b-a, c+a]$. Then, for this partition

$$(f * g)(x_i) - (f * g)(x_{i-1}) = \int_{-\infty}^{\infty} g(t)\, [f(x_i - t) - f(x_{i-1} - t)]\, dt$$

and

$$\sum_{i=1}^{n} |(f * g)(x_i) - (f * g)(x_{i-1})| \leqslant \int_{-\infty}^{\infty} |g(t)| \sum_{i=1}^{n} |f(x_i - t) - f(x_{i-1} - t)|\, dt$$

$$\leqslant \int_{-\infty}^{\infty} |g(t)| \sup_{\mathscr{P}} \sum_{i=1}^{n} |f(x_i - t) - f(x_{i-1} - t)|\, dt$$

$$\leqslant \bigvee_{b-a}^{c+a} f. \quad \blacksquare$$

Let $(\mathscr{L}_a, \|\cdot\|_a)$ be the space of integrable functions with support in $[-a, 1+a]$.

LEMMA 2.3. *Let $f \in \mathscr{L}_a$ and let h be a probability density function with support in $[-a, a]$. Define the bounded linear operator $\mathscr{P}_\lambda \colon \mathscr{L}_a \to \mathscr{L}_a$ by*

$$\mathscr{P}_\lambda = \lambda P_{\tau_a} f + (1-\lambda)(P_{\tau_a} f) * h,$$

where $0 < \lambda < 1$. Then $\mathscr{P}_\lambda$ has a fixed point f_λ with $f_\lambda \geqslant 0$, $\|f_\lambda\|_a = 1$, $\bigvee_{J_a} f_\lambda \leqslant K_a$, for some constant K_a.

Proof. From the proof of Theorem 1 in [6], we know that there exists a constant c_a such that $\limsup_{i\to\infty} \bigvee_{J_a} P_{\tau_a}^i f \leqslant c_a \|f\|_a$. Assume $\|f\|_a = 1$. Then there exists a constant K_a such that for all i

$$\bigvee_{J_a} P_{\tau_a}^i f \leqslant K_a.$$

Now, let

$$S = \left\{ f \in \mathscr{L}_a : f \geqslant 0, \|f\|_a = 1, \bigvee_{J_a} f \leqslant K_a \right\}.$$

S is a convex set, and by Helly's Theorem it is compact in $(\mathscr{L}_a, \|\cdot\|_a)$. Let $f \in S$. Then, clearly $\mathscr{P}_\lambda f \geqslant 0$, $\|\mathscr{P}_\lambda f\|_a = 1$, and by Lemma 2.2,

$$\bigvee_{J_a} \mathscr{P}_\lambda f - \lambda \bigvee_{J_a} P_{\tau_a} f + (1-\lambda) \bigvee_{J_a} \mathscr{P}_{\tau_a} f \leqslant K_a .$$

Hence, $\mathscr{P}_\lambda S \subset S$, and the Markov–Kakutani Theorem establishes the existence of a fixed point $f_\lambda \in S$, i.e., $\mathscr{P}_\lambda f_\lambda = f_\lambda$. ∎

LEMMA 2.4. *Let $\{a_{n,\lambda}(x): n \geqslant 0, 0 < \lambda < 1\}$ be a double sequence of functions in $\mathscr{L}_1$. Let $\{a_{n,\lambda}\}$ converge strongly to $F \in \mathscr{L}_1$ in the Cesaro mean, uniformly in λ, i.e.,*

$$\left\| \frac{1}{n} \sum_{i=0}^{n-1} a_{n,\lambda} - F \right\| \to 0$$

as $n \to \infty$, uniformly in λ. Then

$$\left\| (1-\lambda) \sum_{n=0}^{\infty} \lambda^n a_{n,\lambda} - F \right\| \to 0$$

as $\lambda \to 1^-$.

Proof. A direct consequence of the standard theorem: Cesaro convergence implies Abel convergence [7, Theorem 1.14].

THEOREM 1. *Let $\tau \in \mathscr{C}$ have the unique invariant density f^* and let τ_a be a $\mathscr{C}$-extension to J_a. Let $\{W_\lambda: 0 < \lambda < 1\}$ be a family of random variables with range in $[-a, a]$, where ϕ_λ, the probability density function of W_λ, is continuous. We assume that $\phi_\lambda \Rightarrow \delta_0$ as $\lambda \to 1^-$. Then the invariant probability density functions of the stochastic difference equation*

$$x_{n+1} = \tau_a(x_n) + W_\lambda \tag{4}$$

approach f^* in $\mathscr{L}_1$ as $\lambda \to 1^-$.

Proof. In view of Lemma 2.1, we can approximate $\{\phi_\lambda\}$ by $\{\psi_\lambda\}$, where

$$\psi_\lambda = \lambda\delta_0 + (1-\lambda)\beta_\lambda ,$$

and $\beta_\lambda \in \mathscr{M}(A)$ can be represented by

$$\beta_\lambda = \frac{\phi_\lambda - q_\lambda}{1-\lambda} .$$

Let us now prove the theorem for the approximating family of probability density functions $\{\psi_\lambda : 0 < \lambda < 1\}$. If x_n has probability density function f, then $\tau(x_n) + W_\lambda$ has the approximating probability density function $(P_{\tau_a} f) * \psi_\lambda$. We define $\mathscr{P}_\lambda : \mathscr{L}_a \to \mathscr{L}_a$ by

$$\mathscr{P}_\lambda f = (P_{\tau_a} f) * \psi_\lambda , \tag{5}$$

i.e.,

$$\mathscr{P}_\lambda f = \lambda P_{\tau_a} f + (1 - \lambda)(P_{\tau_a} f) * \beta_\lambda . \tag{6}$$

By Lemma 2.3, we know that $\mathscr{P}_\lambda$ has a fixed point f_λ which is probability density function on J_a and is of bounded variation. Thus

$$f_\lambda = \lambda P_{\tau_a} f_\lambda + (1 - \lambda)(P_{\tau_a} f_\lambda) * \beta_\lambda . \tag{7}$$

On successively substituting f_λ into the right-hand side of (7), we obtain

$$f_\lambda = (1 - \lambda) \sum_{i=0}^{\infty} \lambda^i P_{\tau_a}^i v_\lambda , \tag{8}$$

where $v_\lambda = (P_{\tau_a} f_\lambda) * \beta_\lambda$. Let us write (8) as follows:

$$f_\lambda = (1 - \lambda) \sum_{i=1}^{\infty} \lambda^i P_{\tau_a}^{i-1}(P_{\tau_a} v_\lambda) + (1 - \lambda) v_\lambda . \tag{9}$$

Before proceeding with (9), recall that $P_\tau f^* = f^*$. Since f^* has support in $[0, 1]$,

$$\int_{\tau_a^{-1}(B)} f^*(x)\, dx = \int_{\tau^{-1}(B)} f^*(x)\, dx,$$

and hence $P_{\tau_a} f^* = f^*$. Since f^* is the unique fixed point of τ, it must be the unique fixed point of P_{τ_a} when restricted to $\mathscr{L}_1$. By virtue of [6], this implies that for any $f \in \mathscr{L}_1$

$$\frac{1}{n} \sum_{i=1}^{n-1} P_{\tau_a}^i f \to f^* \tag{10}$$

strongly in $\mathscr{L}_1$ as $n \to \infty$. Since the range of τ_a is the same as that of τ, it follows from formula (6) of [6] that $P_{\tau_a}(\mathscr{L}_a) \subset \mathscr{L}_1$. Hence $P_{\tau_a} v_\lambda \in \mathscr{L}_1$ for each $0 < \lambda < 1$, and (10) can be written as

$$\frac{1}{n} \sum_{i=1}^{n-1} P_\tau^{i-1}(P_{\tau_a} v_\lambda) \to f^* \tag{11}$$

as $n \to \infty$ for each λ. We claim the convergence is in fact uniform in λ. Consider the set $F = \{P_{\tau_a}^i v_\lambda : 0 < \lambda < 1\}_{i=0}^{\infty}$. For any $g \in F$, we have $\| g \|_a \leqslant 1$ and

$\bigvee_{J_a} g \leqslant K_a$, where K_a is the constant defined in Lemma 2.3. Hence, by Helly's Theorem, F is relatively compact in $\mathscr{L}_a$. By Mazur's Theorem, the same is true for

$$\mathscr{B} = \left\{ \frac{1}{n} \sum_{i=1}^{n-1} P_\lambda^{i-1}(P_{\tau_a} V_\lambda) \colon 0 < \lambda < 1 \right\} \subset \mathscr{L}_1 .$$

Let $S_{n,\lambda} = (1/n) \sum_{i=1}^{n-1} P_\tau{}^i v_\lambda$. Now, $\| S_{n,\lambda} - f^* \| \to 0$ as $n \to \infty$ for each $0 < \lambda < 1$. If this convergence were not uniform in λ, then for some $\epsilon > 0$ there would a sequence $\{N, \lambda_N\}$, $N \to \infty$, $\lambda_N \to 1^-$, such that $\| S_{N,\lambda_N} - f^* \| > \epsilon$. Since $\{S_{N,\lambda_N} \colon N \to \infty, \lambda_N \to 1^-\}$ is not a Cauchy sequence, the relative compactness of $\mathscr{B}$ is contradicted. We can now invoke Lemma 2.4, to get

$$s\text{-}\lim_{\lambda \to 1^-} (1 - \lambda) \sum_{i=1}^{\infty} \lambda^i P_\tau^{i-1}(P_{\tau_a} v_\lambda) = f^*.$$

Taking limits on both sides of (9), we get

$$\lim_{\lambda \to 1^-} \| f_\lambda - f^* \| = 0. \tag{12}$$

So far we have shown that the fixed points $\{f_\lambda\}$ of the operators $\{\mathscr{P}_\lambda\}$ converge strongly to f^*. Recall that $\mathscr{P}_\lambda$ was defined for the family $\{\psi_\lambda\} \subset \mathscr{M}(A)$ which approaches $\{\phi_\lambda\}$ as $\lambda \to 1^-$. Let us now consider the operator $\bar{\mathscr{P}}_\lambda \colon \mathscr{L}_a \to \mathscr{L}_a$, defined by

$$\bar{\mathscr{P}}_\lambda f = \lambda P_{\tau_a} f + (1 - \lambda)(P_{\tau_a} f) * \left(\frac{\phi_\lambda - \lambda\delta_0}{1 - \lambda} \right), \tag{13}$$

where we expressed the given probability density function ϕ_λ as

$$\phi_\lambda + \lambda\delta_0 + (1 - \lambda) \left(\frac{\phi_\lambda - \lambda\delta_0}{1 - \lambda} \right).$$

Note that if f is the probability density function of x_n, then $\bar{\mathscr{P}}_\lambda f$ is the exact probability density function of the right-hand side of the stochastic difference equation (4). Let $f \in \mathscr{L}_a$. Then

$$\bar{\mathscr{P}}_\lambda f - \mathscr{P}_\lambda f = \lambda P_{\tau_a} f - (P_{\tau_a} f) * q_\lambda .$$

Recall from Lemma 2.1 that $q_\lambda/\lambda \Rightarrow \delta_0$ as $\lambda \to 1^-$. It follows from the proof of [8, Theorem 9.1] that for each $f \in \mathscr{L}_a$,

$$\left\| P_{\tau_a} f - (P_{\tau_a} f) * \frac{q_\lambda}{\lambda} \right\|_a \to 0 \tag{14}$$

as $\lambda \to 1^-$. Let S be the set defined in the proof of Lemma 2.3. It is relatively compact and $P_{\tau_a}S \subset S$ is relatively compact. From [11, Theorem IV, 8.20], it follows that

$$\lim_{|t|\to 0} \int_{J_a} |F(x-t) - F(x)|\, dx = 0$$

uniformly for $F \in P_{\tau_a}S$. Using this fact in the proof of [8, Theorem 9.1] establishes

$$\lim_{\lambda\to 1^-} \left\| F - F * \frac{q_\lambda}{\lambda} \right\| = 0 \tag{15}$$

uniformly for $F \in P_{\tau_a}S$. Thus, as $\lambda \to 1^-$,

$$\sup_{f\in S} \| \bar{\mathscr{P}}_\lambda f - \mathscr{P}_\lambda f \|_a \to 0. \tag{16}$$

From this it follows that the fixed points of $\bar{\mathscr{P}}_\lambda$ and $\mathscr{P}_\lambda$ in the set S approach each other as $\lambda \to 1^-$. Hence the fixed points of $\bar{\mathscr{P}}_\lambda$ approach f^* as $\lambda \to 1^-$. ∎

Let $G = \text{support } f^*$ and let $G_\lambda = \{G + x: x \in \text{support } \phi_\lambda\}$. We claim G_λ is the only ergodic set for the Markov chain (5). Let H be another ergodic set. $H \not\subset G$ since τ itself would take every $x_0 \in H$ out of G into $G - H$. If $H \cap G_\lambda \subset G_\lambda - G$, then there are points $x_0 \in H \cap G_\lambda$ such that the random variable W_λ takes x_0 into $G_\lambda - G$ with non-zero probability. Thus, the only other possibility is that $H \cap G_\lambda = \phi$. But then there would exist a point $x_0 \in H$ which has probability 0 of entering G. This contradicts the uniqueness of f^*.

Now let p_λ be a fixed point of $\bar{\mathscr{P}}_\lambda$. Then with respect to p_λ, G_λ is an ergodic set. Let $\{x_i\}_{i=0}^{\infty}$ be the solution process of (5). Once x is specified, $\tau(x) + W_\lambda$ is bounded on J_a, Doeblin's condition [13, p. 192] is satisfied. Hence Theorem 6.1 of [13, Chap. V] applies. That is, since G_λ is the only ergodic set, given *any starting* point $x_0 \in J_a$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(x_i) = \int_{J_a} g(\xi)\, p_\lambda(\xi)\, d\xi, \tag{17}$$

where g is integrable with respect to p_λ, for almost all sample sequences. A word of explanation is in order. Whereas in the Birkhoff Ergodic Theorem, convergence is a.e. with respect to the invariant measure on the state space, the convergence in (17) is *everywhere* with respect to that measure, but a.e. with respect to the measure on the sequence space, the space of all possible orbits. Perhaps this important distinction can be made clearer by considering a simple example. Let $0 < \theta < 1$ be the probability of getting heads in a flip of a coin. Let ω denote an experiment, i.e., an infinite sequence of outcomes of Bernoulli trials, and let $(S_n(\omega))/n$ denote the relative frequency of heads in n tosses associated with the experiment ω, the record of outcomes of an infinite sequence of coin tossings. Then, as is well known,

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} = \theta \tag{18}$$

for almost every ω. The exceptional set $\mathscr{A}$ arises from the fact that if a coin is tossed indefinitely, it is conceivable, for example, that heads would occur every time. Such an event never happens in practice since its probability is equal to $\lim_{n\to\infty} \theta^n = 0$. Similarly, for the stochastic difference equation (5), it is conceivable that $W_\lambda = 0$ every time a value of the random variable is chosen; such an occurrence would, of course, render (17) incorrect. But the probability of this happening is zero. The null set $\mathscr{A}$ is different from the null set $\mathscr{K}$ of the invariant measure on the state space for which the Birkhoff Ergodic Theorem fails. For example, for piecewise linear transformations, $\mathscr{K}$ consists of all the rationals. This is, indeed, prohibitively restrictive, whereas $\mathscr{A}$ is nothing more than a theoretical nuisance. Therein lies the difference between the a.e. (sequence space) convergence in (17) and the a.e. (state space) convergence of the Birkhoff Ergodic Theorem.

We now state the final result.

THEOREM 2. *Let $\tau \in \mathscr{C}$ have the unique invariant density f^*. Given any $\epsilon > 0$, we can find $0 < \lambda < 1^-$ and close to 1^- such that the solution process $\{x_i\}_{i=0}^{\infty}$ of the stochastic difference equation*

$$x_{n+1} = \tau_a(x_n) + W_\lambda$$

satisfies

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_B(x_i) = \int_B p_\lambda(\xi)\, d\xi, \tag{19}$$

for all starting points $x_0 \in [0, 1]$, and for almost all sequences, where B is a measurable set of $[0, 1]$, and p_λ satisfies

$$\| p_\lambda - f^* \| \leqslant \epsilon.$$

Proof. Theorem 1.

Remarks. (1) Implementing the left-hand side of (19) produces only an approximation to f^*; it, however, holds for all starting points. Hence, the random perturbation has produced a completely predictable situation out of one—its deterministic counterpart—which was chaotic.

(2) If $\tau(0) = \tau(1)$, then there is no need to extend τ to τ_a. We simply use $\bar{\tau}(x) = \tau(x) \bmod (1)$, in place of $\lambda(x)$.

(3) The method of proof for Theorem 1 was inspired by the statement of Theorem 2 in [16].

3. Examples

Example 1. Consider the piecewise-linear transformation $\tau: [1, 5] \to [1, 5]$ defined ny $\tau(1) = 3$, $\tau(2) = 5$, $\tau(3) = 4$, $\tau(4) = 2$, $\tau(5) = 1$. In [12], it is shown that τ is a Markov map and that it has a unique absolutely continuous invariant measure with density π, constant on each of the four subintervals: normalized, $\pi = (2/7, 1/7, 2/7, 2/7)$. Since the slopes of τ are rational, it can be shown [1] that all the rational points in [1, 5] are periodic or periodic after a finite number of iterations. Thus, the rationals cannot be starting points if one hopes to attain π by taking time averages of the orbit $\{\tau^n(x)\}_{n=0}^{\infty}$.

Let us, therefore, consider the stochastic difference equation

$$x_{n+1} = \tau_a(x_n) + \epsilon W, \tag{20}$$

where $W \sim N(0, 1)$ is obtained in the following way. Let U_1 and U_2 be uniform random variables on [0, 1], obtained by using iterates of the transformation $T(x)$ = fractional part of $(\pi + x)^5$ (see [10]). Then, by 6a(3), Chapter 26.8 of [17], we have that

$$X = (-2 \ln U_1)^{1/2} \cos(2\pi U_2)$$

is an $N(0, 1)$ random variable. Now, let $x_0 = 4.6$ be the starting point. This is a period point of order eight for τ, and as such, $\{\tau^n(4.6)\}_{n=0}^{\infty}$ cannot generate π. Letting $\epsilon = 10^{-8}$, an HP-97 programmable calculator was used to obtain the solution orbit $\{x_i\}_{i=0}^{3423}$ of (20), which is entirely in [1, 5]. Hence τ_a is not needed in (20) for this orbit. The following normalized distribution was obtained on the four subintervals:

$$(0.2807, 0.1440, 0.2857, 0.2857).$$

To four decimal places, the true density π is

$$(0.2857, 0.1428, 0.2857, 0.2857).$$

Example 2. Consider the transformation $\tau: [0, 1] \to [0, 1]$ defined by

$$\begin{aligned} \tau(x) &= 2x, & 0 \leqslant x \leqslant \tfrac{1}{2} \\ &= (2-a) - 2(1-a)\,x, & \tfrac{1}{2} \leqslant x \leqslant 1, \end{aligned}$$

where $0 \leqslant a < \frac{1}{2}$. For these values of a, τ has slope greater than one in absolute value. Therefore, there exists an absolutely continuous invariant measure μ invariant under τ. Since τ has only one discontinity, it is unique [9]. Let f_a be its density. It must be the normalized solution of $P_\tau f_a = f_a$, i.e.,

$$
\begin{aligned}
f_a(x) &= 0 & 0 \leqslant x < a \\
&= cf_a(\tfrac{1}{2} + c(1 - x)), & a \leqslant x < 2a \\
&= \frac{1}{2} f_a\left(\frac{x}{2}\right) + cf_a\left(\frac{1}{2} + c(1 - x)\right), & 2a \leqslant x \leqslant 1,
\end{aligned}
$$

where $c = 1/(2(1 - a))$. Obtaining even approximate solutions of this functional equation is a diffcult matter. With the aid of the stochastic difference equation (5), however, we can easily obtain a close approximation to f_a. Let $W \sim N(0, 1)$ as in Example 1, and let $\epsilon = 10^{-9}$. We use

$$x_{n+1} = \tau(x_n) + 10^{-9} W, \tag{21}$$

We do not bother with a $\mathscr{C}$-extension of τ since for ϵ so small, none of the orbits considered left the interval $[0, 1]$.

The interval $[0, 1]$ is subdivided into 20 equal subintervals. It was found that the difference in the proportions of visits between 1000 and 10,000 iterates is insignificant. Hence, we chose $N = 1000$ to be the number of iterates. For the initial value $x_0 = 0.3$, Fig. 1 shows the distribution of the proportion of visits to each subinterval for values of a approaching 0. It is clear that the distributions become more and more uniform as $a \to 0$. Since it is well known that $f_0(x)$ is the uniform density on $[0, 1]$, the numerical results indicate that there is structural stability with respect to the invariant measures, namely

$$\lim_{a \to 0} \sup_{x \in [0,1]} |f_a(x) - f_0(x)| = 0.$$

For values of a close to $\frac{1}{2}$, the solution is more complex. For $a = \frac{1}{2}$, all points in $(0, \frac{1}{2}]$ go into $[\frac{1}{2}, 1]$ and each point in $[\frac{1}{2}, 1]$ is periodic with period two. For $a < \frac{1}{2}$ and close to it, there exists an absolutely continuous invariant measure. Hence this example provides an oppurtunity to follow the transition from ergodic measures to a discrete measure which is not ergodic. Figure 2 shows that at some a_0, $0.37 < a_0 < 0.4$ (actually $0.37 < a_0 < 0.38$), a gap appears in the density f_{a_0}. As $a \to \frac{1}{2}^-$, this gap widens, and the resulting two-lobed densities approach the measure $\frac{1}{2}\delta_{1/2} + \frac{1}{2}\delta_1$, which characterizes the periodic orbit $\{\frac{1}{2}, 1\}$.

Final Remarks. (1) There resides in the literature the unproven notion that very long orbits behave in a statistically predictable manner. It is clear that

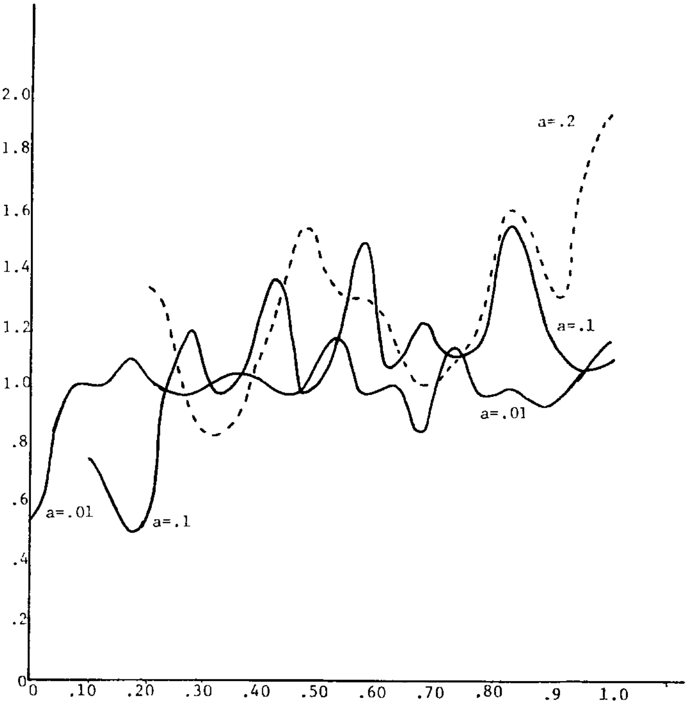

FIG. 1. $\epsilon = 10^{-a}$, $N = 1000$, $X_0 = 0.3$, a approaching 0.

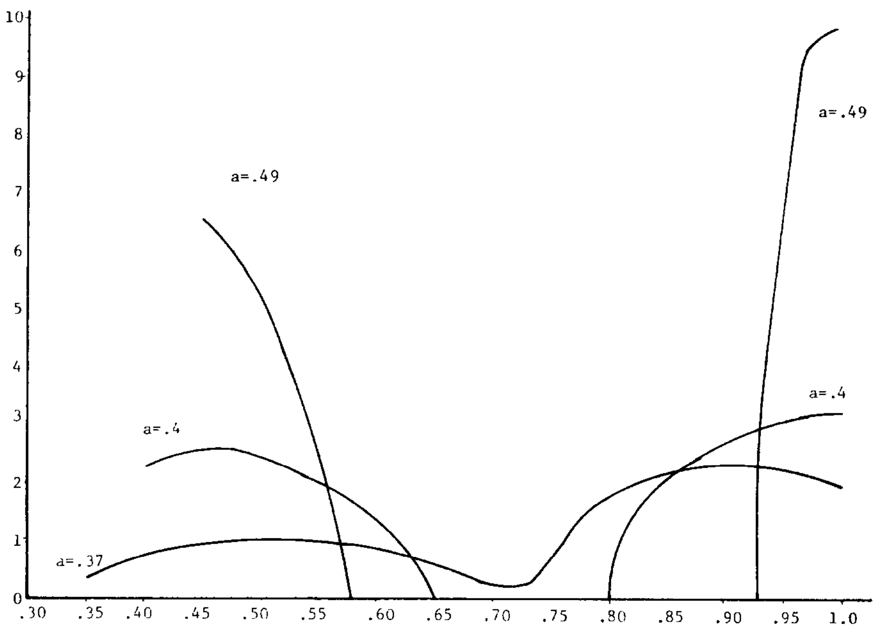

FIG. 2. $\epsilon = 10^{-a}$, $N = 1000$, $X_0 = 0.3$, a approaching 0.5.

in expremely long orbits, the number of points in the orbit exceeds the capacity of the calculator. If the calculator can handle N digits, then 10^N is the largest cycle it will admit. For larger cycles, the orbit will overflow and introduce truncation errors. If these truncation errors are viewed as a kind of random perturbation on a cyclic orbit of admissible order, then the foregoing theory would account for the expected limiting behaviour obtained from long orbits.

(2) On page 465 of [18], Professor May writes: "What seems called for is some effective stochastic description of the deterministic dynamics". This note is perhaps a small step in that direction.

Acknowledgments

The author is grateful to Prof. M. Kanter for a cntical reading of the manuscript and to Prof. A. Lasota for a number of helpful discussions.

References

1. M. Scarowsky, A. Boyarsky, and H. Proppe, Properties of piecewise-linear chaotic functions, *Nonlinear Analysis, Theory, Methods and Appl.* **4** No. 1, pp. 109–121.
2. T-Y Li, Finite approximation for the Frobenius–Perron operator. A solution to Ulam's conjecture, *J. Approximation Theory* **17** (1976), 177–186.
3. S. Ulam, "Problems in Modern Mathematics," Interscience, New York, 1960.
4. A. Lasota, On the existence of invariant measures for Markov processes, *Ann. Pol. Math.* **28** (1973), 207–211.
5. A. N. Kolmogorov and S. V. Fomin, "Introductory Real Analysis," Dover, New York, 1975.
6. A. Lasota and J. A. Yorke, On the existence of invariant measures for piecewise monotonic transformations, *Trans. Amer. Math. Soc.* **186** (1973), 481–488.
7. R. E. Powell and S. M. Shah, "Summability Theory and Applications," Van Nostrand Reinhold, New York, 1972.
8. R. L. Wheeden and A. Zygmond, "Measure and Integral," Dekker, New York, 1977.
9. T-Y Li and Yorke, Ergodic transformations from an interval into itself, *Trans. Amer. Math. Soc.* **235** (1978), 183–192.
10. T-Y Li and J. A. Yorke, Ergodic maps on [0, 1] and nonlinear pseudo-Random number generators, *Nonlinear Anal.* **2**, No. 4 (1978), 273.
11. N. Dunford and J. T. Schwartz, "Linear Operators, Part I," Wiley, New York, New York, 1967.
12. A. Boyarsky and M. Scarowsky, A class of transformations which have unique absolutely continuous invariant measures, *Trans. Amer. Math. Soc.* **255** (1979), 243–262.
13. J. L. Doob, "Stochastic processes," Wiley, New York, 1953.
14. T-Y Li and J. A. Yorke, Period three implies chaos, *Amer. Math. Monthly* **82** (1975), 985–992.
15. P. Kloeden, Chaotic difference equations are dense, *J. Austral. Math. Soc.* **15** (1976), 371–375.

16. A. LASOTA, "A Solution of Olam's Conjecture on the Existence of Invariant Measures and its Applications, Dynamical Systems: An International Symposium" (L. Cesari, J. Hale, and J. P. Lasalle, Eds.), Academic Press, New York, 1976.
17. M. ABROMOWITZ AND I. A. STEGUN, "Handbook of Mathematical Functions," National Bureau of Standards, Applied Mathematics Series, 55, 1965.
18. R. M. MAY, Simple mathematical models with very complicated dynamics, *Nature* **261** (1976), 459–467.